

# GMD-Based Hybrid Precoding For Millimeter-Wave Massive MIMO Systems

Tian Xie, Linglong Dai, Xinyu Gao, Muhammad Zeeshan Shakir, and Zhaocheng Wang

**Abstract**—Hybrid precoding can significantly reduce the number of required radio frequency (RF) chains and relieve the huge energy consumption in mmWave massive MIMO systems, thus attracting much interests from academic and industry. However, most existing hybrid precoding schemes are based on singular value decomposition (SVD). Due to the very different sub-channel signal-to-noise ratios (SNRs) after SVD, complicated bit allocations is usually required to match the sub-channel SNRs. To solve this problem, we propose a geometric mean decomposition (GMD)-based hybrid precoding scheme to avoid the complicated bit allocation. Its basic idea is to seek a pair of analog and digital precoding matrices that are sufficiently close to the optimal unconstrained GMD precoding matrix. Specifically, we design the analog (digital) precoding matrix while keeping the digital (analog) precoding matrix fixed. Further, the principle of basis pursuit is utilized in the design of analog precoding matrix, while we obtain the digital precoding matrix by projecting the GMD operation on the digital precoding matrix. Simulation results verify that the proposed GMD-based hybrid precoding scheme outperforms conventional SVD-based hybrid precoding schemes and achieves much better bit error rate (BER) performance with low complexity.

**Index Terms**—Massive MIMO, mmWave communications, hybrid precoding, geometric mean decomposition.

## I. INTRODUCTION

Due to the vast unlicensed spectrum in millimeter-wave (mmWave) band (30-300 GHz), mmWave communications can offer huge spectrum resource [1]. On the other hand, massive MIMO can provide enough array gain through a large number of antennas (e.g., 256 antennas) to compensate for the severe attenuation of mmWave signals [2]. Thus, the combination of mmWave communication and massive MIMO (mmWave massive MIMO) is a promising technology for the future 5G wireless communications, since it can significantly increase the system throughput [3]. However, the conventional fully digital precoding structure, where one dedicated radio frequency (RF) chain is required for each antenna, is too energy-consumptive due to the large number of RF chains (e.g., 256 RF chains for 256 antennas) in mmWave massive MIMO systems. To

address this problem, a hybrid analog/digital precoding structure has been proposed [4], which decomposes the fully digital precoding matrix into one low-dimensional digital precoding matrix and one high-dimensional analog precoding realized through analog circuits (e.g., phase shifter networks) [4]. In this way, the hybrid precoding structure only needs a small number of RF chains to achieve the near-optimal performance [4].

How to obtain the optimal precoding matrix is the key issue for hybrid precoding. In [5], the authors formulated the hybrid precoding design problem as a sparse reconstruction problem, and proposed to use the orthogonal matching pursuit (OMP) algorithm to obtain the analog precoding matrix. [6] proposed a low-complexity version of the algorithm in [5] based on the local search method, while [7] derived a lower bound on the numbers of RF chains to perfectly achieve the performance of fully digital precoding by using the hybrid precoding structures. The basic ideas of these works above are seeking a pair of analog and digital precoding matrices that are sufficiently close to the right singular matrix from the singular value decomposition (SVD) of the channel matrix (which are called SVD-based hybrid precoding in this paper). SVD-based hybrid precoding can achieve the capacity-approaching performance with the help of water-filling power allocation. Nevertheless, due to the very different SNRs for each sub-channel in SVD-based precoding, complicated bit allocations (i.e., allocating different modulation and coding schemes (MCS) on different sub-channels) are needed to match the SNRs of different sub-channels, which involve high or even unaffordable coding/decoding complexity in practice [8].

In this paper, we propose a hybrid precoding scheme for mmWave massive MIMO systems based on the geometric mean decomposition (GMD) to avoid the complicated bit allocations. Specifically, different from the conventional SVD-based hybrid precoding, we propose to treat the right semi-unitary matrix in GMD as the optimal unconstrained precoding matrix, which can convert the mmWave massive MIMO channel into identical sub-channels and therefore naturally do not require bit allocations any more. Furthermore, to efficiently find a solution to the GMD-based hybrid precoding design problem, which is difficult to be solved, we use a decoupled optimization method to design the analog and digital precoding matrix. Particularly, we solve the analog precoding matrix design problem via the principle of basis pursuit inspired by [5], while we obtain the digital precoding matrix by projecting the GMD operation on the digital precoding matrix. Simulation results verify that the proposed GMD-based hybrid precoding scheme outperforms the conventional SVD-based

T. Xie, L. Dai, X. Gao, and Zhaocheng Wang are with the Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Electronic Engineering, Beijing 100084, China (e-mail: razor9321@163.com, daill@mail.tsinghua.edu.cn, gxy1231992@sina.com, and zcwang@mail.tsinghua.edu.cn).

Muhammad Zeeshan Shakir is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. (e-mail: muhammad.shakir@sce.carleton.ca).

This work was supported in part by the International Science & Technology Cooperation Program of China (Grant No. 2015DFG12760), the National Natural Science Foundation of China (Grant Nos. 61571270 and 61271266), the Beijing Natural Science Foundation (Grant No. 4142027), and the Foundation of Shenzhen government.

hybrid precoding schemes and achieves much better bit error rate (BER) performance with low complexity.

*Notation:* Lower-case and upper-case boldface letters  $\mathbf{a}$  and  $\mathbf{A}$  denote a vector and a matrix.  $\text{tr}\{\mathbf{A}\}$ ,  $\{\mathbf{A}\}^T$ ,  $\{\mathbf{A}\}^H$ ,  $\|\mathbf{A}\|_F$ , and  $\{\mathbf{A}\}_{i,j}$  denote the trace, transpose, conjugate transpose, Frobenius norm, and the element in the  $i$ th row and the  $j$ th column of  $\mathbf{A}$ , respectively.  $a := b$  is the in-place update operation, where  $a$  is replaced by  $b$ .

## II. SYSTEM MODEL

### A. Hybrid analog/digital precoding

We consider a typical mmWave massive MIMO system with hybrid precoding structure, where the base station (BS) with  $N_t$  transmit antennas sends  $N_s$  independent data streams to the user with  $N_r$  receiving antennas. Furthermore, we assume that the BS and the user have  $N_t^{\text{RF}}$  and  $N_r^{\text{RF}}$  RF chains, respectively, which satisfy  $N_s \leq N_t^{\text{RF}} \leq N_t$  and  $N_s \leq N_r^{\text{RF}} \leq N_r$  [5]. In the hybrid precoding structure, as shown in Fig. 1, the hybrid precoding matrix  $\mathbf{P} \in \mathbb{C}^{N_t \times N_s}$  at the BS can be written as the product of two parts: the first part is a low-dimension digital precoding matrix  $\mathbf{P}_D \in \mathbb{C}^{N_t^{\text{RF}} \times N_s}$ ; the second part is a high-dimension analog precoding matrix  $\mathbf{P}_A \in \mathbb{C}^{N_t \times N_t^{\text{RF}}}$ , i.e.,  $\mathbf{P} = \mathbf{P}_A \mathbf{P}_D$ , where  $\mathbb{C}$  is the set of complex numbers. Note that the above remarks on precoding matrix can also be applied to the combining matrix. Thus, the transmitted signal vector  $\mathbf{x}$  is

$$\mathbf{x} = \mathbf{P}\mathbf{s} = \mathbf{P}_A \mathbf{P}_D \mathbf{s}, \quad (1)$$

where  $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$  denotes the source signal vector. To meet the constraint of transmitting power, we bound the transmit power at the BS as  $\text{tr}\{\mathbf{P}\mathbf{P}^H\} \leq N_s$ . After receiving the signal vector, the user utilizes a hybrid combining matrix  $\mathbf{W}$  to combine the signal vector:

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho} \mathbf{W}^H \mathbf{H} \mathbf{x} + \mathbf{W}^H \mathbf{n} \\ &= \sqrt{\rho} \mathbf{W}_B^H \mathbf{W}_A^H \mathbf{H} \mathbf{P}_A \mathbf{P}_D \mathbf{s} + \mathbf{W}_D^H \mathbf{W}_A^H \mathbf{n}, \end{aligned} \quad (2)$$

where  $\rho$  is the average received power,  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  denotes the channel matrix between BS and the user, and  $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$  is the additive white Gaussian noise (AWGN) vector at the user following the circularly symmetric complex Gaussian distribution with zero mean vector and covariance matrix  $\sigma^2 \mathbf{I}_{N_r}$ , i.e.,  $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$ , with  $\sigma^2$  the noise covariance, and  $\mathbf{I}_{N_r}$  the  $N_r$  by  $N_r$  identity matrix. Since the analog precoder is realized through analog phase shifter network after up-converters, all elements of  $\mathbf{P}_A$  should have the same amplitude:

$$|\{\mathbf{P}_A\}_{i,j}| = \frac{1}{\sqrt{N_t}}, \quad (3)$$

where  $|\cdot|$  denotes the modulus of a complex number.

### B. Channel Model

To capture the limited spatial selectivity or scattering characteristics in mmWave massive MIMO channel, we adopt a widely used Saleh-Valenzuela (SV) model in this paper [5], [9], where the channel matrix  $\mathbf{H}$  can be expressed as:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \sum_i^L \beta_i \mathbf{a}_r(\varphi_i^r) \mathbf{a}_t(\varphi_i^t), \quad (4)$$

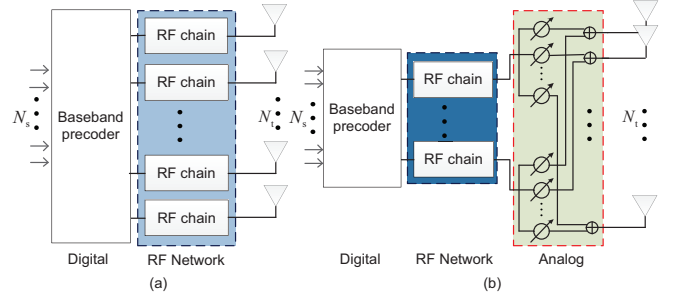


Fig. 1. Comparison between two typical precoding structures: (a) The fully digital precoding; (b) The hybrid analog/digital precoding.

where  $L$  denotes the number of paths,  $\beta_i$  is the complex gain of the  $i$ th path,  $\mathbf{a}_r(\varphi_i^r)$  and  $\mathbf{a}_t(\varphi_i^t)$  are the array response vectors at the user and the BS, respectively, where  $\varphi_i^r$  denotes the angle of arrival (AoA) at the user, and  $\varphi_i^t$  is the angle of departure (AoD) at the BS. For the simple uniform linear line (ULA) antenna array of  $N$  elements, the array response vector is

$$\mathbf{a}_{\text{ULA}}(\varphi) = \sqrt{\frac{1}{N}} [1, e^{jk d \sin(\varphi)}, \dots, e^{j(N-1)k d \sin(\varphi)}]^T, \quad (5)$$

where  $k = \frac{2\pi}{\lambda}$  with  $\lambda$  the wavelength, and  $d$  denotes the antenna spacing. Because of the limited spatial scattering in mmWave propagation, the mmWave massive MIMO channel  $\mathbf{H}$  is low-rank [10]. As a result, we can leverage a finite number of RF chains to achieve the near-optimal throughput [10].

## III. PROPOSED GMD-BASED HYBRID PRECODING

In this section, we first review the fully digital precoding schemes. Next, we propose the GMD-based hybrid precoding.

### A. Fully digital SVD-and GMD-based precoding

In this subsection, we first briefly review the fully digital precoding schemes in mmWave massive MIMO systems. Denote the SVD of the channel matrix  $\mathbf{H}$  by

$$\begin{aligned} \mathbf{H} &= \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \\ &= [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} \\ &= \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^H + \mathbf{U}_2 \mathbf{\Sigma}_2 \mathbf{V}_2^H, \end{aligned} \quad (6)$$

where  $\mathbf{U}_1 \in \mathbb{C}^{N_r \times N_s}$  and  $\mathbf{V}_1 \in \mathbb{C}^{N_t \times N_s}$  are semi-unitary matrices containing the left  $N_s$  columns of unitary matrices  $\mathbf{U} \in \mathbb{C}^{N_r \times N_r}$  and  $\mathbf{V} \in \mathbb{C}^{N_t \times N_t}$ , respectively, and  $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \dots, \sigma_{N_s})$  is a diagonal matrix with the largest  $N_s$  singular values of  $\mathbf{H}$  in decreasing order. With fully digital SVD-based precoding, i.e.,  $\mathbf{P} = \mathbf{V}_1$  and  $\mathbf{W} = \mathbf{U}_1^H$ , the MIMO channels are converted into  $N_s$  parallel sub-channels, of which the sub-channel gains are  $\sigma_1, \dots, \sigma_{N_s}$ , i.e.,

$$\begin{aligned} \mathbf{y} &= \sqrt{\rho} \mathbf{W}^H \mathbf{H} \mathbf{P} \mathbf{s} + \mathbf{W}^H \mathbf{n} \\ &= \sqrt{\rho} \mathbf{U}_1^H \mathbf{H} \mathbf{V}_1 \mathbf{s} + \mathbf{U}_1^H \mathbf{n} = \sqrt{\rho} \mathbf{\Sigma}_1 \mathbf{s} + \mathbf{U}_1^H \mathbf{n}. \end{aligned} \quad (7)$$

Due to the limited spatial scattering in mmWave propagation, the singular values of channel matrix  $\mathbf{H}$  varies a lot [10], which results in the very different SNRs among different sub-channels after water-filling power allocation<sup>1</sup> as shown in Fig.

<sup>1</sup>Note that the water-filling algorithm allocates less power on the sub-channel with lower channel gain, which worsens the SNR for this sub-channel.

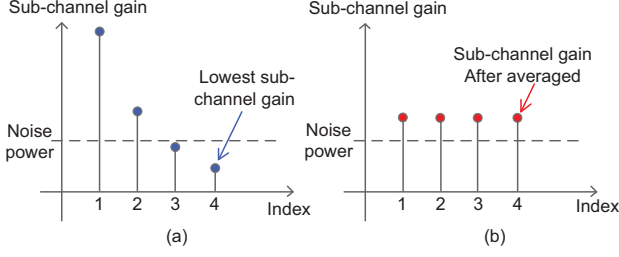


Fig. 2. Intuitive illustration of sub-channel gains: (a) Conventional SVD-based hybrid precoding; (b) Proposed GMD-based hybrid precoding.

2 (a). Consequently, if we use the same MCS for all sub-channels, the BER performance for the MIMO system will be determined by the sub-channel with the lowest SNR. On the other hand, if we want to guarantee the similar BER performance among all sub-channels, careful bit allocations are needed, which can bring the unaffordable coding/decoding complexity in mmWave massive MIMO systems [8].

To solve this problem, GMD has been proposed [11], where the channel matrix  $\mathbf{H}$  is decomposed as:

$$\mathbf{H} = \mathbf{G}\mathbf{R}\mathbf{Q}^H = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & * \\ \mathbf{0} & \mathbf{R}_2 \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^H \\ \mathbf{Q}_2^H \end{bmatrix}, \quad (8)$$

where  $\mathbf{G}_1 \in \mathbb{C}^{N_r \times N_s}$  and  $\mathbf{Q}_1 \in \mathbb{C}^{N_t \times N_s}$  are semi-unitary matrices containing the left  $N_s$  columns of unitary matrices  $\mathbf{G} \in \mathbb{C}^{N_r \times N_r}$  and  $\mathbf{Q} \in \mathbb{C}^{N_t \times N_t}$ ,  $\mathbf{R}_1$  is an upper triangular matrix with identical diagonal elements<sup>2</sup>, i.e., the geometric mean of singular values  $r_{i,i} = (\sigma_1 \sigma_2 \cdots \sigma_{N_s})^{\frac{1}{N_s}} = \bar{r}$ ,  $\forall i$ , and  $*$  denotes an arbitrary matrix that we do not care. Note that the diagonal element in  $\mathbf{R}_1$  is the geometric mean of the first  $N_s$  singular values of  $\mathbf{H}$ , since there are only  $N_s$  data streams. The fully digital GMD-based precoding employs  $\mathbf{Q}_1$  as the precoding matrix and  $\mathbf{G}_1^H$  as the combining matrix, where we can transform the effective channel matrix in (2) into an upper triangular matrix:

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{W}^H \mathbf{H} \mathbf{P} \mathbf{s} + \mathbf{W}^H \mathbf{n} \\ &= \mathbf{G}_1^H \mathbf{H} \mathbf{Q}_1 \mathbf{s} + \mathbf{G}_1^H \mathbf{n} = \mathbf{R}_1 \mathbf{s} + \mathbf{G}_1^H \mathbf{n}. \end{aligned} \quad (9)$$

Then, by utilizing successive interference cancellation (SIC) at the receiver [11], we can acquire  $N_s$  sub-channels with equal sub-channel gain  $r_{ii}$  as shown in Fig. 2 (b), which means that we can avoid the bit allocation caused by different SNRs for different sub-channels in precoding schemes based on SVD.

### B. GMD-based hybrid precoding

In this subsection, we turn to discuss the GMD-based hybrid precoding. According to (1) and (4), we can formulate the hybrid precoding problem as follow<sup>3</sup> [5]:

$$\begin{aligned} (\mathbf{Q}_A^{\text{opt}}, \mathbf{Q}_D^{\text{opt}}) &= \arg \min_{\mathbf{Q}_A, \mathbf{Q}_D} \|\mathbf{Q}_1 - \mathbf{Q}_A \mathbf{Q}_D\|_F, \\ \text{s.t. } & \|\{\mathbf{Q}_A\}_{i,j}\| = \frac{1}{\sqrt{N_t}}, \\ & \text{tr}(\mathbf{Q}_A \mathbf{Q}_D \mathbf{Q}_D^H \mathbf{Q}_A^H) \leq N_s. \end{aligned} \quad (10)$$

<sup>2</sup>Here, we use the simple form  $r_{i,j}$  to denote the element in the  $i$ th row and  $j$ th row of  $\mathbf{R}_1$  for simplicity.

<sup>3</sup>To distinguish the GMD-based hybrid precoding from the SVD-based precoding, we use  $\mathbf{Q}_1$ ,  $\mathbf{Q}_A$ , and  $\mathbf{Q}_D$  to denote the unconstrained GMD-based precoding matrix, the GMD-based analog precoding matrix, and the GMD-based digital precoding matrix, respectively.

This problem formulation can be intuitively understood as seeking a pair of  $\mathbf{Q}_A \mathbf{Q}_D$  sufficiently “close” to the optimal unconstrained precoding matrix  $\mathbf{Q}_1$ .

The following **Lemma 1** shows some insights about how to efficiently solve (10).

**Lemma 1.** *The GMD of  $\mathbf{H}$  can be expressed as a rotation of SVD as following*

$$\mathbf{Q}_1 = \mathbf{V}_1 \mathbf{S}_R; \quad \mathbf{G}_1 = \mathbf{U}_1 \mathbf{S}_L; \quad \mathbf{R}_1 = \mathbf{S}_L^T \mathbf{\Sigma}_1 \mathbf{S}_R \quad (11)$$

where  $\mathbf{S}_R \in \mathbb{C}^{N_s \times N_s}$  and  $\mathbf{S}_L \in \mathbb{C}^{N_s \times N_s}$  are unitary matrices dependent on  $\mathbf{\Sigma}_1$ .

*Proof:* According to [12], the implementation algorithm of GMD based on SVD can be briefly summarized as follow:

1) *Step1:* Initialize  $\mathbf{R}_1 = \mathbf{\Sigma}_1$ ,  $\mathbf{Q}_1 = \mathbf{V}_1$ ,  $\mathbf{G}_1 = \mathbf{U}_1$ , and compute the geometric mean value  $\bar{r} = (\sigma_1 \sigma_2 \cdots \sigma_{N_s})^{\frac{1}{N_s}}$ .

2) *Step2:* At stage  $i$ , where  $i$  varies from 1 to  $N_s - 1$ , first check the  $i$ th diagonal element of  $\mathbf{R}_1$ , i.e.,  $r_{i,i}$ : If  $r_{i,i} \geq \bar{r}$ , find certain element  $r_{p,p}$  such that  $r_{p,p} \leq \bar{r}$ , where  $p > i$ ; Otherwise, find certain element  $r_{p,p}$  such that  $r_{p,p} \geq \bar{r}$ , where  $p > i$ . Then, swap  $r_{i+1,i+1}$  and  $r_{p,p}$ , as well as the corresponding  $(i+1)$ th and  $p$ th column in  $\mathbf{Q}_1$  and  $\mathbf{G}_1$  by

$$\mathbf{R}_1 := \{\mathbf{M}^{(i)}\}^T \mathbf{R}_1 \mathbf{M}^{(i)}, \quad \mathbf{Q}_1 := \mathbf{Q}_1 \mathbf{M}^{(i)}, \quad \mathbf{G}_1 := \mathbf{G}_1 \mathbf{M}^{(i)}, \quad (12)$$

where  $\mathbf{M}^{(i)}$  is the corresponding permutation matrix.

3) *Step3:* Construct two Givens matrices  $\mathbf{N}_L^{(i)}$  and  $\mathbf{N}_R^{(i)}$  from  $\mathbf{I}_{N_s}$  by replacing the sub-matrix containing the four elements  $\{\mathbf{N}_L^{(i)}\}_{i,i}$ ,  $\{\mathbf{N}_L^{(i)}\}_{i+1,i}$ ,  $\{\mathbf{N}_L^{(i)}\}_{i,i+1}$ , and  $\{\mathbf{N}_L^{(i)}\}_{i+1,i+1}$  with the following  $2 \times 2$  matrices:

$$\mathbf{\Theta}_L^{(i)} = \frac{1}{\bar{r}} \begin{bmatrix} cr_{i,i} & sr_{i+1,i+1} \\ -sr_{i+1,i+1} & cr_{i,i} \end{bmatrix}, \quad \mathbf{\Theta}_R^{(i)} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \quad (13)$$

respectively, where

$$c = \sqrt{\frac{\bar{r}^2 - r_{i+1,i+1}^2}{r_{i,i}^2 - r_{i+1,i+1}^2}}, \quad s = \sqrt{1 - c^2}. \quad (14)$$

Then, update  $\mathbf{R}_1$ ,  $\mathbf{Q}_1$ , and  $\mathbf{G}_1$  as

$$\mathbf{R}_1 := \mathbf{N}_L^{(i)} \mathbf{R}_1 \mathbf{N}_R^{(i)}, \quad \mathbf{Q}_1 := \mathbf{Q}_1 \{\mathbf{N}_L^{(i)}\}^T, \quad \mathbf{G}_1 := \mathbf{G}_1 \mathbf{N}_R^{(i)}. \quad (15)$$

One can verify that

$$\mathbf{\Theta}_L^{(i)} \begin{bmatrix} r_{i,i} & 0 \\ 0 & r_{i+1,i+1} \end{bmatrix} \mathbf{\Theta}_R^{(i)} = \begin{bmatrix} \bar{r} & * \\ 0 & \frac{r_{i,i} r_{i+1,i+1}}{\bar{r}} \end{bmatrix}, \quad (16)$$

which means that by  $\mathbf{N}_L^{(i)} \mathbf{R}_1 \mathbf{N}_R^{(i)}$ , we can make  $r_{i,i} = \bar{r}$ , while keeping other elements unchanged.

4) *Step4:* Update  $i := i + 1$  and go back to Step 2 until  $i$  equals  $N_s - 1$ .

Combining (15) and the fact that  $\mathbf{M}^{(i)}$ ,  $\mathbf{N}_L^{(i)}$ , and  $\mathbf{N}_R^{(i)}$  are all unitary matrices, we can confirm **Lemma 1** by letting  $\mathbf{S}_L = \mathbf{M}^{(1)} \mathbf{N}_R^{(1)} \mathbf{M}^{(2)} \mathbf{N}_R^{(2)} \cdots \mathbf{M}^{(N_s)} \mathbf{N}_R^{(N_s)}$  and  $\mathbf{S}_R = \mathbf{M}^{(1)} \{\mathbf{N}_L^{(1)}\}^T \mathbf{M}^{(2)} \{\mathbf{N}_L^{(2)}\}^T \cdots \mathbf{M}^{(N_s)} \{\mathbf{N}_L^{(N_s)}\}^T$ . ■

According to **Lemma 1**, the objective function in (10) can be written as

$$\begin{aligned} \|\mathbf{Q}_1 - \mathbf{Q}_A \mathbf{Q}_D\|_F &= \|\mathbf{V}_1 \mathbf{S}_R - \mathbf{Q}_A \mathbf{Q}_D\|_F \\ &\stackrel{(a)}{=} \|\mathbf{V}_1 - \mathbf{Q}_A \mathbf{Q}_D \{\mathbf{S}_R\}^H\|_F \\ &= \|\mathbf{V}_1 - \mathbf{Q}_A \tilde{\mathbf{Q}}_D\|_F, \end{aligned} \quad (17)$$



where (a) holds because the Frobenius norm is invariant under rotations, and  $\tilde{\mathbf{Q}}_D = \mathbf{Q}_D \{\mathbf{S}_R\}^H$ . From (17), we can observe that the optimal unconstrained precoding matrix in (10) will be  $\mathbf{V}_1$ , if we rotate  $\mathbf{Q}_D$  as  $\tilde{\mathbf{Q}}_D = \mathbf{Q}_D \{\mathbf{S}_R\}^H$ . So we reformulate (10) as

$$\begin{aligned} (\mathbf{Q}_A^{\text{opt}}, \mathbf{Q}_D^{\text{opt}}) &= \arg \min_{\mathbf{Q}_A, \tilde{\mathbf{Q}}_D} \|\mathbf{V}_1 - \mathbf{Q}_A \tilde{\mathbf{Q}}_D\|_F, \\ \text{s.t. } |\{\mathbf{Q}_A\}_{i,j}| &= \frac{1}{\sqrt{N_t}}, \\ \text{tr}(\mathbf{Q}_A \tilde{\mathbf{Q}}_D \mathbf{Q}_D^H \tilde{\mathbf{Q}}_A^H) &\leq N_s. \end{aligned} \quad (18)$$

However, solving (18) is also highly complicated, since  $\mathbf{Q}_A$  and  $\tilde{\mathbf{Q}}_D$  are coupled, and the elemental-wise constraints  $|\{\mathbf{Q}_A\}_{i,j}| = 1/\sqrt{N_t}$  is non-convex [5]. To this end, we propose to solve to (18) by decoupling the design  $\mathbf{Q}_A$  and  $\mathbf{Q}_D$ , i.e., treating  $\mathbf{Q}_D$  as a fixed matrix while designing  $\mathbf{Q}_A$ , and vice versa. Note that the main difficult lies in the design of the analog precoding matrix  $\mathbf{Q}_A$  due to a non-convex constraint on  $\mathbf{Q}_A$ . To effectively design the analog precoding matrix, inspired by [5], we leverage the array response vectors  $\mathbf{a}_t(\varphi_i^t), \forall i$  as the columns of  $\mathbf{Q}_A$ , which can be explained as follow. Firstly,  $\mathbf{a}_t(\varphi_i^t), \forall i$  has constant modulus elements, which satisfy the constraint  $|\{\mathbf{Q}_A\}_{i,j}| = 1/\sqrt{N_t}$ . Secondly, the columns of  $\mathbf{V}_1$  form an orthogonal basis of the channel's row space, while from (4), the array response vectors  $\mathbf{a}_t(\varphi_i^t), \forall i$  also form an orthogonal basis of the channel's row space. As a result,  $\mathbf{V}_1$  can be expressed as a linear combination of  $\mathbf{a}_t(\varphi_i^t), \forall i$ . In addition, the hybrid precoding matrix  $\mathbf{Q}_A \tilde{\mathbf{Q}}_D$  can be also seen as a linear combination of  $\mathbf{Q}_A$ . Since we want to approximate  $\mathbf{V}_1$  via  $\mathbf{Q}_A \tilde{\mathbf{Q}}_D$ , it is reasonable to find the "best"  $N_t^{\text{RF}}$  array response vectors from  $\mathbf{a}_t(\varphi_i^t), \forall i$  as the columns of  $\mathbf{Q}_A$ . So the corresponding analog precoding design problem becomes

$$\begin{aligned} \mathbf{T} &= \arg \min_{\mathbf{T}} \|\mathbf{V}_1 - \mathbf{A}_t \mathbf{T} \tilde{\mathbf{Q}}_D\|_F, \\ \text{s.t. } \|\text{diag}(\mathbf{T} \mathbf{T}^H)\|_0 &= N_t^{\text{RF}}, \\ \text{tr}(\mathbf{A}_t \mathbf{T} \tilde{\mathbf{Q}}_D \tilde{\mathbf{Q}}_D^H \mathbf{T}^H \mathbf{A}_t^H) &\leq N_s, \end{aligned} \quad (19)$$

where  $\mathbf{A}_t = [\mathbf{a}_t(\varphi_1^t), \mathbf{a}_t(\varphi_2^t), \dots, \mathbf{a}_t(\varphi_L^t)]$  is an  $N_t \times L$  matrix, and  $\mathbf{T}$  is a selecting matrix with  $N_t^{\text{RF}}$  non-zero rows. Problem (19) is a sparse reconstruction problem, which can be effectively solved via the principle of basis pursuit [5]. After we determine the analog precoding matrix, the digital precoding matrix design problem is a Frobenius norm minimization problem:

$$\begin{aligned} \mathbf{Q}_D^{\text{opt}} &= \arg \min_{\tilde{\mathbf{Q}}_D} \|\mathbf{V}_1 - \mathbf{Q}_A \tilde{\mathbf{Q}}_D\|_F, \\ \text{s.t. } \text{tr}(\mathbf{Q}_A \tilde{\mathbf{Q}}_D \mathbf{Q}_D^H \tilde{\mathbf{Q}}_A^H) &\leq N_s. \end{aligned} \quad (20)$$

The optimal solution to (20) has a least square form [9]

$$\tilde{\mathbf{Q}}_D = \{\mathbf{Q}_A\}^\dagger \mathbf{V}_1. \quad (21)$$

Furthermore, we have

$$\mathbf{Q}_D = \tilde{\mathbf{Q}}_D \mathbf{S}_R = \{\mathbf{Q}_A\}^\dagger \mathbf{V}_1 \mathbf{S}_R. \quad (22)$$

which can be seen as projecting the GMD operation on the digital precoding matrix  $\mathbf{Q}_D$  to (20).

#### Require: $\mathbf{H}$ ;

1) Perform SVD of channel:  $[\mathbf{U}_1 \ \Sigma_1 \ \mathbf{V}_1] = \text{SVD}(\mathbf{H})$ ;

2) Initialize  $\mathbf{Q}_A = \text{Empty matrix}$ , and  $\mathbf{Q}_{\text{res}} = \mathbf{V}_1$

**For**  $i \leq N_t^{\text{RF}}$

3)  $\Phi = \mathbf{A}_t^H \mathbf{Q}_{\text{res}}$ ;

4)  $k = \arg \max_{l=1, \dots, L} (\Phi \Phi^H)_{l,l}$ ;

5)  $\mathbf{Q}_A = [\mathbf{Q}_A | \mathbf{A}_t^{(k)}]$ ;

6)  $\mathbf{Q}_D = \mathbf{Q}_A^\dagger \mathbf{V}_1$ ;

7)  $\mathbf{Q}_{\text{res}} = \frac{\mathbf{V}_1 - \mathbf{Q}_A \mathbf{Q}_D}{\|\mathbf{V}_1 - \mathbf{Q}_A \mathbf{Q}_D\|_F}$ ;

**End for**

8) Generate  $\mathbf{S}_R$  based on the algorithm in **Lemma 1**, and  $\mathbf{Q}_D = \mathbf{Q}_D \mathbf{S}_R$ ;

9) Normalize  $\mathbf{Q}_D = \sqrt{N_s} \frac{\mathbf{Q}_D}{\|\mathbf{Q}_A \mathbf{Q}_D\|_F}$ ;

10) Return  $\mathbf{Q}_A$  and  $\mathbf{Q}_D$ ;

**Algorithm 1:** The proposed GMD-based hybrid precoding

The overall algorithm for GMD-based hybrid precoding is summarized in **Algorithm 1**. **Algorithm 1** can be divided into two parts in general. The first part includes step 1 to step 7, which performs conventional spatially sparse hybrid precoding [5] based on the optimal unconstrained precoding matrix  $\mathbf{V}_1$ . The second part includes step 8 and step 9, which performs the GMD operation on the digital precoding matrix  $\mathbf{Q}_D$  and normalizes the effective precoding matrix to meet the transmitting power constraints. Note that in step 8, we do not need to explicitly compute  $\mathbf{S}_R$ , instead, we just apply the corresponding permutation and multiplication operations in each stage on  $\mathbf{Q}_D$ . Note that the complexity of implementing GMD given SVD, i.e., Step 8, is only  $\mathcal{O}((N_s + N_t)N_s)$  [12], since only a  $N_s$  by  $N_s$  rotation matrix  $\mathbf{S}_R$  is multiplied to  $\mathbf{Q}_D$ , while the complexity of hybrid precoding algorithm, i.e., Step 1-7, is  $\mathcal{O}(N_{\text{RF}}^2 N_t N_s)$ , since we need to compute the pseudo-inverse matrix of  $\mathbf{Q}_A$ , which indicates that the GMD-based hybrid precoding will only incur small extra complexity compared with conventional SVD-based hybrid precoding.

## IV. SIMULATION RESULTS

In this section, we will evaluate the performance of the proposed GMD-based hybrid for mmWave massive MIMO communications. Consider typical mmWave massive MIMO systems where  $N_t = 128$ -and  $N_t = 256$ -element ULAs with the antenna spacing  $d = \lambda/2$  are utilized at the BS, while  $N_r = 16$ -element ULA with also the antenna spacing  $d = \lambda/2$  antennas is utilized at the user. The BS and the user both adopt  $N_t^{\text{RF}} = N_r^{\text{RF}} = 4$  RF chains, and the carrier frequency is 28GHz. For the channel model, we utilize the SV model, where the number of path is set to  $L = 4$ , the complex gain of each path follows the distribution  $\mathcal{CN}(0, 1)$ , and the AoAs and AoDs are uniformly distributed in  $[-\pi/2, \pi/2]$ . Finally, we use the water-filling power allocation scheme at the BS [5], while the modulation scheme is 16QAM, which is utilized in all sub-channels for both SVD and GMD-based precoding schemes.

Fig. 3 shows the BER performance of the proposed GMD-based hybrid precoding in a  $128 \times 16$  mmWave massive MIMO system, where we can observe that the proposed GMD-based precoding schemes (including fully digital GMD-based

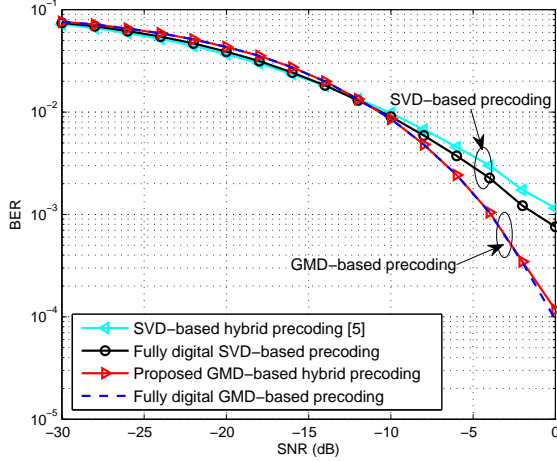


Fig. 3. BER performance comparison of the proposed GMD-based hybrid precoding in a  $128 \times 16$  mmWave massive MIMO system.

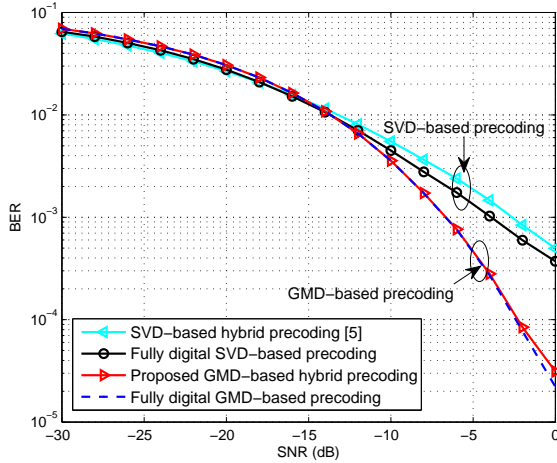


Fig. 4. BER performance comparison of the proposed GMD-based hybrid precoding in a  $256 \times 16$  mmWave massive MIMO system.

precoding and GMD-based hybrid precoding) can achieve a better BER performance than the conventional SVD-based precoding (including fully digital SVD-based precoding and SVD-based hybrid precoding [5]). This is because the GMD-based precoding converts the mmWave MIMO channel into multiple sub-channels with the identical SNR, which naturally avoids the severe BER performance degradation over sub-channels with very lowest SNR for conventional SVD-based precoding. Specifically, given the same BER performance, e.g.,  $\text{BER} = 10^{-3}$ , the required SNR for GMD-based precoding schemes is about  $-4\text{dB}$ , which is lower than that of conventional SVD-based precoding. Furthermore, we can also see that the performance gap between the fully digital GMD-based precoding and the proposed GMD-based hybrid precoding is negligible, which implies that the proposed GMD-based hybrid precoding is able to well approximate the fully digital GMD-based precoding.

Fig. 4 is the BER performance of the proposed GMD-based hybrid precoding in a  $256 \times 16$  mmWave massive MIMO system. In Fig. 4, the BER of the proposed GMD-based precoding and the conventional SVD-based precoding both improves, since more antennas at the BS provide more

array gains. However, the proposed GMD-based precoding still outperforms the conventional SVD-based precoding. The required SNR in GMD-based hybrid precoding to achieve  $\text{BER} = 10^{-3}$  is about  $2.5\text{dB}$  and  $3.5\text{dB}$  lower than that in the fully digital SVD-based precoding and the SVD-based hybrid precoding [5]. In addition, the proposed GMD-based hybrid precoding can still perform close to the fully digital GMD-based precoding with negligible gap.

## V. CONCLUSIONS

In this paper, we propose a GMD-based hybrid precoding scheme for mmWave massive MIMO systems to avoid the complicated bit allocations in the conventional SVD-based hybrid precoding. With the help of SIC at the receiver, we can convert the mmWave MIMO channel into multiple identical sub-channels, which naturally do not need bit allocations. Specifically, we decouple the design of the analog and digital precoding matrix. The analog precoding matrix is designed using the principle of basis pursuit, while we project the GMD operation on the digital precoding matrix to obtain the digital precoding matrix. Simulation results verify that the proposed GMD-based hybrid precoding scheme outperforms conventional SVD-based hybrid precoding schemes and is able to achieve much better BER performance.

## REFERENCES

- [1] T. Bai, A. Alkhateeb, and R. Heath, "Coverage and capacity of millimeter-wave cellular networks," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 70–77, Sep. 2014.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [3] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.
- [4] S. Han, C.-L. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid precoding analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, Jan. 2015.
- [5] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [6] C. Rusu, R. Mendez-Rial, N. Gonzalez-Prelcicy, and R. W. Heath, "Low complexity hybrid sparse precoding and combining in millimeter wave MIMO systems," in *Proc. 2015 IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 1340–1345.
- [7] E. Zhang and C. Huang, "On achieving optimal rate of digital precoder by RF-baseband codesign for MIMO systems," in *Proc. 80th IEEE Veh. Technol. Conf. (VTC Fall)*, Sep. 2014, pp. 1–5.
- [8] C.-L. Chao, S.-H. Tsai, and T.-Y. Hsu, "Bit allocation schemes for mimo equal gain precoding," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1345–1350, Mar. 2011.
- [9] X. Gao, L. Dai, S. Han, and R. W. Heath Jr, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, May. 2016.
- [10] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran, and M. Z. Shakir, "Mmwave massive-MIMO-based wireless backhaul for the 5G ultra-dense network," *IEEE Wireless Communications*, vol. 22, no. 5, pp. 13–21, Oct. 2015.
- [11] Y. Jiang, W. W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra and Its Applications*, vol. 396, pp. 373–384, Feb. 2005.
- [12] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Tran. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.